JOINT WORKING GROUP ON METHODOLOGICAL ISSUES IN CLINICAL TRIALS IN RADIOLOGICAL SCREENING AND RELATED COMPUTER MODELING

Sponsored by
U.S. Public Health Service's Office on Women's Health
and the National Cancer Institute

Co-Sponsored by the American College of Radiology
January 25-26, 1999

Summary Recommendations

Introduction to Scientific Program   **William C. Black, M.D.**
Dartmouth-Hitchcock Medical Center

Constantine Gatsonis, Ph.D.
Brown University

New Methodological Research Areas

Methods of Analysis

Estimation Problems

- Distribution-free methods for estimating sojourn time distribution of preclinical sojourn time
- Estimating sensitivities of one or more detection modalities with estimates of distribution-free sojourn times
- Estimate sensitivities and sojourn time distributions as a function of age, other factors
- Nonrandomized studies
- Group randomization
- Noncompliance

New Experimental Design for Planning Early Detection Trials

- Benefit for all groups
- Evaluating more than one modality
- Methods for early assessment of trials, e.g., surrogate markers, stage, etc.
- Methods for inferring natural time history of disease, e.g., sequence of events and time between events
- Assemble central database of early detection trials, which would be publicly available

Public Health Programs

- **Planning Programs:** More than one modality, ages for exams, high-risk populations
- **Evaluation** of potential benefits of public health programs
- **Recommendations** for future exams as function of history

  - o Models of early detection process

  - o Tests having continuous outcomes

  - o   Incorporation of more than one test modality

Session 1: Empirical Studies of Cancer Screening: Study Designs, Choice of Endpoint, Strengths and Limitations

Les M. Irwig, M.B.B.Ch., Ph.D.
University of Sydney, Australia

William E. Barlow, Ph.D.
University of Washington

A. Previous approaches to making decisions about screening have not paid enough attention to the harms (downsides, non-monetary costs) of screening.
   1. Decision analytic models should include harms.
   2. Models should weigh benefits and harms to obtain net effectiveness before considering financial costs.
   3. Methods should be developed to better inform individuals and the community about the absolute benefit and harms of screening.
   4. Methods should be developed to elicit community and individual preferences to ensure that people who are screened are those who consider the benefit: harm tradeoff favorable.
   5. The concepts and application of community and individual informed consent need to be developed and applied.

B. We do not always need results on efficacy of screening to make decisions about screening.
   1. Develop decision analytic models before doing any studies and decide what studies are necessary on the basis of where the data are most needed.

C. New technologies need to be rapidly evaluated.
   1. Initial studies should be randomized trials to assess whether treatment of disease that is detected early improves person-centered outcomes (survival and quality of life).
   2. Procedures need to be developed to ensure that promising new technologies are identified early and that structures exist to rapidly implement individual or cluster randomized trials started with minimal delay and red tape. Trials should be designed to ensure maximum statistical power at minimal cost and time delay. Methods need to be developed to improve the efficiency of trials and address ethical concerns.
   3. Studies of new technologies for early detection of diseases (for which early treatment has been shown to be effective by trials) can use surrogate measurements. Best studies are either RCTs with test performance or surrogates as the outcome or studies comparing test performance characteristics on the same individuals.

D. Surrogate assessment using data from past trials may be inappropriate because the screening procedure and availability of surrogate information may have altered since that time.

    1. Further research is needed on the extent to which surrogate measures can correctly predict survival and quality of life. This is best done by ensuring that future randomized trials collect full information on surrogates and store serum and other patient material for later analysis.

    2. Methods for assessing surrogates in observational studies need to be developed.

    3. Methods for choosing surrogates may be different if the objective is to evaluate new technologies or implement quality control in local settings.

E. Research on new experimental designs of planning early detection trials to evaluate benefit of early detection of disease.

In particular:

    1. Experimental designs which offer potential benefits to all participants

    2. Taking account of the special features of early detection trials:
- sample size
- number of exams
- interval between exams
- sensitivity and specificity of exams
- prevalence and incidence of disease
- compliance

    3. Optimal design on these special features

    4. Design issues on unbalanced design for control and study groups

    5. Design issues on different screening exams given to control and study participants

Session 2:   Model-Based Evaluations of Cancer Screening: Survey of Current Approaches, Methodological Issues and Challenges Integrating Modeling with Empirical Studies

Gerrit J. van Oortmarssen, Ph.D.
Erasmus University Rotterdam

Stuart G. Baker, Sc.D.
National Cancer Institute

A. Role of Models

   1. Direct relation to screening trials
- design
- early indicators of outcome
- refine estimates (noncompliance, dilution)
- meta-analyses

   2. Observational data
- monitoring of screening
- nonrandomized designs (bias)

   3. Extrapolation/optimization
- screening interval, age range,....
- cost-effectiveness

B. Types of Models
   1. Deep/mechanistic (explicit natural history)
- tumor growth
- discrete state
  - microsimulation models: flexible

   2. Empirical models, specific topics
- noncompliance, dilution
- periodic screening evaluation
- meta-analysis including heterogeneity

 C. Quality Of Models - Requirements
   1.   Use best data available
   2. Plausible assumptions*
   3. Proper validation
   4. Clear documentation (transparent, complete)
   5. Sensitivity/uncertainty analyses

*Find most parsimonious model that fits data.

D. Future Directions
1. Uncertainty/sensitivity analysis
2. Additional sources of data
3. Combine features of models
4. Quality of life
5. New designs of trials

Session 3: Modeling Disease Development and Progression

William C. Black, M.D.
Dartmouth-Hitchcock Medical Center

Sylvia K. Plevritis, Ph.D.
Stanford University

A. NAT HX cannot be directly observed.

B. NAT HX tightly constrained by ING. mortality, and autopsy findings.

C. Integrate above with information on tumor biology.
   1. Prob of met as fn of 1 tumor char
   2. Prob that in situ de progress to inv ca

D. Properly integrate NAT HX model with accuracy of screening process (stage specific).

E. Properly integrate above with effectiveness of early diagnosis.

F. To use model to help design empirical studies, including RCTs and OBS studies.

G. Recognize tension between simplicity and complexity and role for both.

Session 4: Incorporating Variations in Test Accuracy

Craig A. Beam, Ph.D.
Northwestern University Medical School

Constantine Gatsonis, Ph.D.
Brown University

Studies have shown that diagnostic accuracy of an imaging technology can vary significantly by
- reader
- subjects
- characteristics of the technology.

These sources of variations can significantly impact the effectiveness of radiological screening programs and need to be incorporated in the design and evaluation of screening interpretations.

Methodological questions need to be addressed to answer the question of *how* these sources should be measured and then used.

In radiological screening, variations in test accuracy in no small part arise from variability among humans in their interpretation of, and decision-making about, the information provided by imaging technology. Other sources of variation that can impact the accuracy of tests can come about with decisions made for further workup, etc. Therefore, *variations* in test accuracy are reflective of the *human/actors* inherent in the use of diagnostic technology.

Methodological questions that need to be addressed at present are as follows:

A. Measurements
1. How best to measure the accuracy of human diagnosticians
2. How relevant (generalizable) "test-based" evaluations are to actual field performance
3. How best to decompose reader variability into intrinsic *diagnostic accuracy* vs. *situational performance*
4. Need further development and evaluation of retrospective studies and prospective studies
    - Retrospective studies
        - "generalizability" of "reader studies"
        - "Failure analysis" of systems
        - how applicable and useful?
    - Prospective studies
        - substantial sample size
        - gold standard
5. Need to measure variability at important levels of consolidation (Health Systems)
    - Practice-level
    - Health care system
    - States
    - Socioeconomic strata, etc.

B. Evaluation
1. How should variations be incorporated in models of screening?
   - Average value
   - Distribution of accuracies - distribution of effectiveness
   - Use "summary ROC" curves generated by meta-analysis (but need to recognize that problems from primary empirical studies can be inherited)

2. How should basic fact of sensitivity/specificity tradeoff be incorporated?

C. Proscriptive
1. How can we put this "diversity" to best use?
2. Develop computer models of distributed diagnostic systems

Session 6: Modeling Heterogeneity and Uncertainty: Heterogeneity Across Individuals, Uncertainty in Model Assumptions

Donald A. Berry, Ph.D.
Duke University

Giovanni Parmigiani, Ph.D

A. Heterogeneity
1. Individual heterogeneity
- Risk prediction models
- Does efficacy depend on covariates, e.g., mammography and BRCA?
- Biological heterogeneity

2. Study heterogeneity
- Modeling sources of heterogeneity

B. Uncertainty in comprehensive decision models

1. Access to primary data
2. Validation, transparency, availability of models
3. Computational strategies
4. Integration versus posterior probabilistic sensitivity analysis versus scenario-based sensitivity analysis

C. General point
1. Trials/models interactions
- Decision trials so that we can develop better models, e.g., surrogate endpoint sojourn time estimation, interactions between covariates and efficacy

Session 7: Incorporating Utilities and Costs

Alvin I. Mushlin, M.D., Sc.M.
University of Rochester Medical Center

Scott D. Ramsey, M.D., Ph.D.
University of Washington

A. Quality-adjusted outcomes (HR QofL and H UTBL) (not just survival) are important in models.

B. Include outcomes specific to diagnostic/screening tests (i.e., not just improving therapeutic decisions). For example:
- Reassurance and anxiety reduction to patients
- Reducing uncertainty to physicians

C. Above best assessed after people are provided with accurate information.

D. Include costs in models and use concepts of C/E analysis in optimization decisions.

E. Censoring should be accounted for when estimating costs.

F. Cost should include all relevant costs;
- Program costs
- Screening procedures
- Patient travel
- Value of time lost during screening
- Cost of treating side effects
- Economic savings from cancers averted

G. Databases and RCTs such as SEER-Medicare can be useful sources of costs.

Session 8: Screening and Decision-Making: Optimizing Screening Strategies

Ping Hu, SC.D.
Brown University

J.D.F. Habbema, Ph.D.
Erasmus University Rotterdam

A. 1. Decision-making concerns planning, monitoring, adapting, and evaluation of screening (including followup).
   2. Status (validation; possibilities) of models used should be clarified.
   3. Situation for which calculations are intended (time, place, population).
   4. Standards for CEA (panel of CEA...) should be followed, in particular, incremental CER, societal cost perspective, discount percentage, LY, and QALY.
   5. In general: standardization.

B. 1. Not only final CER results, but also intermediate steps (insight).
   2. Sensitivity analysis and evt. uncertainty analysis.
   3. Quality of life: measurement and/or sensitivity analysis or qualitative discussion.
   4. Due attention to communication of results.
   5. Decision-makers/organizers/etc... involved in the study.

C. Attention to high/familial/genetic risk groups (only incidence changed?)
D. Comparative modeling studies?

Session 9: Counseling and Clinical Decision-Making

Karia M. Kerlikowske, M.D.
University of California at San Francisco

Steven Woloshin, M.D., M.S.
Veterans Administration Hospital

A. Informing women about the benefits and harms of screening mammography is feasible.

B. Internet is a good modality, (http://mammography.ucsf.edu/inform/index.cfm)

C. Exposure to a presentation of harms and benefits together was associated with a more accurate understanding than presentations "in isolation."

D. Outcomes that matter to people should help set the research agenda.

E. To promote informed decision-making, we need to effectively communicate to people about the likely consequences of their decisions.

F. Researchers/experts may view consequences of screening differently than the public.